Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 70 (2015) 9 – 15

Procedia Computer Science

4th International Conference on Eco-friendly Computing and Communication Systems, ICECCS 2015

An Adaptive Method of PCA for Minimization of Classification Error Using Naïve Bayes Classifier

Devesh Kumar[a]*, Ravinder Singh[b], Abhishek Kumar[c], Nagesh Sharma[d]

[a]*Govt. Engineering. College, Ajmer, India*
[b]*Govt. Engineering College, Ajmer, India*
[c]*MDS University, Ajmer, India*
[d]*Bhagwant University,Ajmer, India*

Abstract

This paper focuses on comparative study of calculation of classification error with classical PCA technique and our adaptive method. Frontal face image database with uniform lightening condition and color images with different orientations have taken in order to get better accuracy in the result. Conventional PCA algorithm is applied for dimensional reduction and calculating the classification error. The result generated from this technique is being compared with the result generated from our adaptive method Naïve Bayes Classifier. We have used Naïve Bayes Classifier for calculating the classification error of each feature vector instead of considering K largest Eigen-value as in PCA. A covariance matrix is arranged by considering the feature vector having the lowest K minimum error. We have applied the adaptive technique on 626 colored facial images with uniform illumination conditions and varying poses and 545 frontal facial images with uniform background to get the better accuracy.

Peer-review under responsibility of the Organizing Committee of ICECCS 2015

Keywords: Principal Component Analysis, Naïve Bayes Classifier, Classification error, Euclidian distance, Minimum Error, Eigen Value.

1. Introduction

Principal component analysis (PCA) came in existence in 1901 by Karl Pearson. PCA is a Dimension reduction procedure and useful for compact relevant information when obtained data have some redundancy. This will result into reduction of variables into smaller number of variables which are called Principal Components which will account for the most of the variance in the observed variable. PCA is a way of identifying patterns in data, and hence

* Corresponding author. Tel.: +919694082008
*E-mail address:*devsharma1507@gmail.com

Peer-review under responsibility of the Organizing Committee of ICECCS 2015
doi:10.1016/j.procs.2015.10.018

Expressing the data in such a way as to highlight their similarities and differences of the face images. Each face image in the database set can be represented exactly in terms of a linear combination of the Eigen faces[2]. This technique allows the system to represent the necessary information for comparing the faces using the little information once the mathematical representation accomplished which it is need to have a lot of faces to be store. On the other hand, it suffers a bit from the fact that facial image have to be normalized that meaning they all have to be the same size and the eyes, nose, and mouth in the sample images must be lined up before the PCA applied. The Principal Component Analysis is one of the most efficient and accurate techniques that have been used in gender and face[3,9] recognition. The PCA algorithm reduce the large dimensionality of the data we are having to the smaller dimensionality of independent feature space which are needed to describe the data. The jobs of PCA are to do data compression, prediction, redundancy removal, feature extraction[2], etc. We have extracted principal directions of the covariance ellipse as done in PCA, the proposed algorithm has not blindly taken the Eigen vectors corresponding to k largest values. Instead, we have transformed the data vectors into the new n–dimensional (n is dimension of old input space) vector space spanned by the Eigen vectors of the covariance matrix of the input data and then one attribute at a time is taken to perform classification[1]. Then, attributes corresponding to k-largest accuracy measures are chosen. A Naïve Bayes Classifier proceeds by assuming that the absence (or presence) of a peculiar feature of a class is distinct to the absence (or presence) of any other feature. Relying on the explicit nature of the probability model, we can train Naïve Bayes Classifier very proficiently in a supervised learning environment. Naive Bayes classifiers are exceptionally adaptable, obliging various parameters linear in the number of variables (predictors/features) in a learning issue. Most extreme probability preparing should be possible by figuring a shut structure expression, which takes linear time, instead of by lavish iterative close estimation as used for some different sorts of classifiers.

2. Principal Component Analysis

2.1 Mathematical Approach of PCA

We are considering 2-D image for working so first of all it will get converted into 1-D vector by concatenating rows and columns. Suppose we have M vectors each of size N (rows * columns) representing a set of sampled images[1].

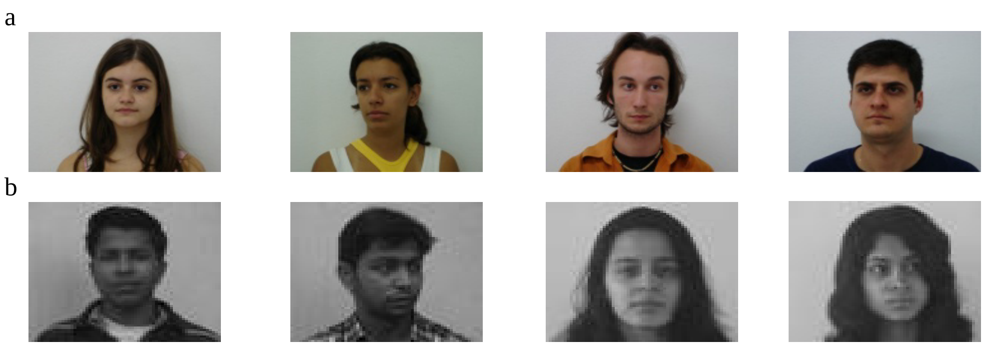

Fig.1. Sample Images with different orientations (a) Color facial Images (b) Black and White facial images.

Let ‘p_j’ represent the values of the pixels.

$$X_i = [\, p_1, p_2, p_3, p_4, \ldots, p_N] \; ; \; i = 1,2,3,4,\ldots M. \tag{1}$$

Then the images are mean centred when we subtract the mean image from each image vector[1,6]. Let us suppose m as the mean image:

$$m= (1/M)* (\Sigma X_i) \tag{2}$$

Let W_i be the mean centred image:

$$W_i= X_i– m \tag{3}$$

Ultimately we have to find the values of e_i's which have the largest possible projection onto each of the w_i's. The purpose is to get M orthogonal vectors e_i for which the quantity[1].

$$\lambda_i = (1/M)\ \Sigma\ (e_i^T * w_n)^2 \tag{4}$$

Is normalized with the orthogonality constraint:

$$e_l^T*e_k = \delta_{lk} \tag{5}$$

The values of e_i's and λ_i's are calculated from the Eigen vectors and the Eigen values of the covariance matrix[1]:

$$C = W*W^T \tag{6}$$

W is a matrix formed by the column vectors w_i places side by side. The size of the covariance matrix is enormous (N*N). It is not possible to solve for eigenvectors directly. In mathematics, there are areas where one needs to find the numbers λ and the vectors v that satisfy the equation where A is the square matrix[1]:

$$Av=\lambda v \tag{7}$$

Any λ satisfying the above equation is the Eigen value of A. The vector v is called the eigenvector of A. The Eigen values and eigenvectors are obtained by solving the equation:

$$[A - \lambda I] = 0 \tag{8}$$

For every λ, we have to normalize the corresponding Eigen-vectors after calculating them. These eigenvectors are then arranged in ascending order which gives us the final KL Transform matrix. The covariance matrix of the final transformed image will have the Eigen-values as their diagonal elements. Moreover, the mean f the final image would be zero[6].

3. Naïve Bayes Classifier

An upper hand of using naïve Bayes Classifier is that it needs a little amount of training data to judge the parameters essential for classification. As independent variables are supposed, rather than evaluating entire covariance matrix only the variances of the variables for each class need to be evaluated. The Naïve Bayes algorithm, a classification algorithm which follows Bayes rules assumes the attributes $X_1 \ldots X_n$ are all conditionally independent of each other[9], given Y. The importance of this assumption is that it dramatically explicates the illustration of $P(X|Y)$, and the issue of evaluating it from the training data. Consider the case where $X = \langle X_1, X_2 \rangle$. In this case,

$$\begin{aligned} P(X|Y) &= P(X_1, X_2|Y) \\ &= P(X_1|X_2, Y)P(X_2|Y) \\ &= P(X_1|Y)P(X_2|Y) \end{aligned}$$

Here the second line adopts a universal property of probabilities, and the third line adopts conditional independence.

In general, when X holds n attributes which are conditionally independent of each other given Y, we have,

$$P(X_1 \ldots X_n \mid Y) = \prod_{i=1}^{n} P(X_i|Y) \tag{9}$$

Observe if Y and the X_i are Boolean variables then we require just 2n parameters to define $P(X_i = x_{ik} \mid Y=y_j)$ for the necessary i, j, k. This is an impressive diminution compared to the $2(2^n - 1)$ parameters needed to describe P(X |Y) if we make no conditional independence supposition[9].

For deriving Naive Bayes algorithm, suppose that Y is any discrete-valued variable, and the attributes $X_1 \ldots X_n$ are any real or discrete valued attributes. Our point is to prepare a classifier that will yield probability distribution for each new instance X that we request that it classify over feasible values of Y. According to Bayes rule, the expression that Y will take for the probability on its k^{th} feasible value, is

$$P(Y = y_k|X_1 \ldots X_n) = \frac{P(Y = y_k)P(X_1 \ldots X_n|Y = y_k)}{\sum_j P(Y = y_j)P(X_1 \ldots X_n|Y = y_j)} \tag{10}$$

Where the sum is taken over every single feasible values y_i of Y. now let us suppose that for given Y, X_i are conditionally independent. Therefore, we can use equation (9) for rewriting this as,

$$P(Y = y_k|X_1 \ldots X_n) = \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_j)} \tag{11}$$

For the Naive Bayes classifier, Equation (11) is the fundamental equation. Given a new instance $X^{new} =<X_1 \ldots X_n>$, this equation simplifies how to calculate the probability that Y will take on any given value, given the observed attribute values of X^{new} and given the distributions P(Y) and $P(X_i|Y)$ approximated from the training data. If we are interested only in the most probable value of Y, then we have the Naive Bayes classification rule:

$$Y \leftarrow \arg\max_{y_k} \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_j)} \tag{12}$$

Which simplifies to the following as denominator does not depends on y_k.

$$Y \leftarrow \arg\max_{y_k} P(Y = y_k)\prod_i P(X_i|Y = y_k) \tag{13}$$

4. Problem with Conventional PCA

In traditional PCA the larger eigenvectors comparing with the larger Eigen values are chosen as principal components, which make the fluctuation of data least. However, there are still some questions of PCA should be addressed as the following. Firstly, the thought of PCA is right from image compression perspective; keeping the biggest non-zero principal components implies that we keep the majority of the energy (information) of that image by anticipating into lower dimension subspace. However, with a view for pattern classifications this argument may not be true. The principle reason is that, in pattern classification we might want to locate a set of projection vectors that can give the most astounding separation between different classes. In this manner, picking the largest principal components as the premise for dimensionality reduction may not be ideal. Also, PCA is statistical type algorithm and unsupervised to extract features in input data since it doesn't utilize the class information[5] of input data. In specific case, the extracted features may not be fitted for classifications. In this way, the principal components are not generally helpful for classification since they are not the most segregating feature. A portion of the little principal components may have preferred execution of classification over the chose large principal components. As we have discussed in the earlier segment emphasizing that PCA is no doubt a revolutionary tool as far as dimension reduction is concerned but when it comes in pattern recognition sometimes it mixes the data pixel[1,7].

Our Proposed Solution-In addition to the traditional PCA which sometimes mixes the data pixel that will further increase the classification error as final consequence. The concept of minimization of approximation error in traditional PCA[4] is get applicable in our work. When the data points get extracted from the image it should be hold their optimal position so that better classification result get occur, we have commuted some concept in traditional PCA in a way that we have considered k- minimum feature vector extracted from the image so what We have done basically is have choose a feature vector projected along k- minimum error, instead of considering the k- largest Eigen value corresponding feature vector[1].

Now the main and the basic thing in doing the above alteration in conventional PCA is to rely upon some tool or method, which will calculate the error with great degree of accuracy so we have chosen Naive Bayes Classifier.

5. Proposed Algorithm with Alteration in Conventional PCA

Iterative_PCA$_{min\ error}$ (D^M, e_t)

1. Calculate covariance of $D^M \rightarrow D^{mx}$. (D^M is data matrix)
2. Perform PCA on D^m to obtain $D^M{}_P$ with PCA-scores arranged in ascending order of the most significant Eigen value projected along k- minimum classification error.
3. Get the pc-score with most significant Eigen value=$e_{min\ error}$ with count being k (say).
4. M_i= MEAN OF normalized pc- scores.(pc= principal component)
5. M_i=zero vector=0 m_{serror}=most significant unit Eigen-vector with respect to k- minimum error.
   While d (M_i, m_{serror})> d (M_{i+1}, m_{serror}) $m_{serror} \leftarrow$ most Significant unit Eigen vector e^t =(e^t - Δe^t).
6. $D^m \leftarrow D^m$ – (data records corresponding to k – minimum error- pc score).
7. If D^m is not empty repeat steps 1 to 5 with new D^M and e^t.

The algorithm we have contemplated is based on traditional PCA algorithm considering the assumption that data points are independent of each other otherwise it would be capacious and complex calculations, so what alteration we have done is basically is to project the feature vector projected corresponding to k-minimum error unlike the traditional PCA consider the feature vector projected along k-largest Eigen value , the motivation of this work is not to recognize the face basically, In the conventional PCA we consider the vector projected along k largest Eigen value like-

λ_{max1} , λ_{max2}, λ_{max3} …… $\lambda_{max\ n}$

In proposed method we have considered the Eigen value corresponding to k-minimum error like-

$\lambda_{min\ error1}$, $\lambda_{min\ error2}$, $\lambda_{min\ error3}$ ……………….. $\lambda_{min\ error\ n}$.

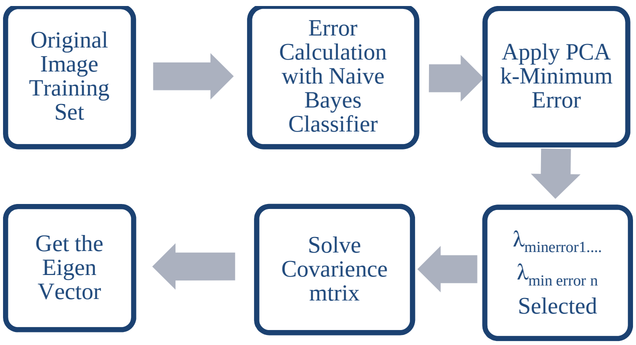

Fig.2. Flowchart of altered approach

6. Comparative study of analysis and experimental results

We have done our work on two different types of image databases; one is color images with different orientations of faces, another one is database consisting of frontal facial black and white images. We have taken database consisting of 626 color image with varying poses and 545 frontal facial black and white images.

We have considered 22 images of each database separately at one time to calculate classification error by both approaches.

Table 1. Comparison between Classification Error calculated by PCA and Naïve Bayes Classifier on 22 colored and 22 black and white images

Final Experimental Results showing Minimization of Classification Error	In PCA(%)	In Naïve Bayes Classifier (%)
1. In case of Frontal Facial Black and white images	0	0
2. In case of colored images with different orientations	0	0

So in order to get the accurate result we have taken 545 images of frontal facial black and white images and 626 color images with different orientations separately once worked on each one altogether at one time and calculated the classification error of the vector corresponding to k- largest value and every time we have got the result biased toward our method as far as error minimization is concerned. What we have observed that in conventional approach λ is not increasing or decreasing in particular pattern. If there would be the outcome like "greater the Eigen value corresponding to the smallest classification error then our proposed method would certainly be diminished reliability wise, but we have to be continuous with the assumption that all data points are independent with each other. This very implementation provides a reliable platform to our approach of PCA for Gender recognition[1,7]. we have made that assumption with a view to ignore any complexity and capacious calculation. The result of the implementation is given below, which is illustrating clearly that the classification error corresponding to the Eigen value can be increase or decrease in random pattern i.e. the conventional PCA is not completely reliable as far as classification error is concerned. The point to give this proof through the experimental results is just to provide a reliable platform to our proposed concept as the output generated in the result is showing classification error.

We have implemented our work to get the right classification error percentage by taking 545 images of males and females and then by calculating the classification error of each value of λ going through with column wise. The final result we have got that is 84% with Naïve Bayes Classifier in case of frontal facial black images as compare to the classical PCA technique which has given 70% of classification error.

Table 2. Comparison between Classification Error calculated by PCA and Naïve Bayes Classifier on 626 colored and 545 black and white images

Final Experimental Results showing Minimization of Classification Error	In PCA(%)	In Naïve Bayes Classifier(%)
1. In case of Frontal Facial Black and white images	70	84
2. In case of colored images with different orientations	88	92

The final result we have got is 92% in case of colored images with different orientations with Naïve Bayes Classifier as compare to the classical PCA technique which has given 88% of classification error. We have gone through with bigger image database from IIT Kanpur and MIT, USA in order to get better accuracy.

7. Conclusion and Future Work

After doing a thoroughly review of PCA application in the domain of gender recognition along with the different algorithm scope in this field such as back propagation, Naïve Bayes Classifier, graph matching. Although our main emphasis is on the improvement of conventional PCA in the direction of classification error minimization by calculating the generalized error in the Eigen value or feature vector going further to be used, while doing the review regarding the gender recognition with PCA what we have got that conventional PCA has scope in it to get improved in regarding the classification error. The paper[6] has given a healthy and important clue in this direction. The main drawback of PCA is that it mixes data points at some phase of classification which generally mislead the classification procedure as this process has lots of scope to ignore the error so in order to overcome this very drawback and with an aim to minimize the classification error[10]. The existing work[1] shows the application of conventional PCA in which the feature vector considered corresponding to k-largest value but what our work primarily all about is to minimize the network error rather than to recognize the face. With the improved approach to the conventional method in which We have considered the feature vector corresponding to the k- minimum error in order to reduce the dimensionality[1,7], we have use Naive Bayes Classifier further to calculate the error and in order to achieve better accuracy comparatively. In order to have better classification, low classification error is very important in training sample, so Naive Bayes Classifier will play an important role in error calculation. Unlike conventional PCA we are considering minimum error Eigen value and then we are using Euclidean distance method in order to get the similarity between the faces in training set and test set. We are planning in future to include more images and make our technique to go through artificial neural network, Naive Bayes Classifier, in order to check out its reliability and flexibility and heterogeneous compatibility. This proposed algorithm can also be checked in speech recognition as well to reduce dimensionality.

References:

1. Abhishek Kumar, Deepak Gupta, Karan Rawat "An Advance Approach of PCA for Gender Recognition" ICICES IEEE ISBN-978-1-4673-5788-3, February 2013.
2. H. C¸ Evikalp, M. Neamtu, M.Wilkes, and A. Barkana. Discriminative common vectors for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1):4–13, 2005.
3. Tan X, Chen S, Zhou Z, Zhang F (2006) Face recognition from a single image per person: a survey. Pattern Recognition 39(9):1725–1745.
4. Abhilash Alexander Miranda, Yann-Ael Le Borgne, Gianluca Bontemp September 12, 2007 New Routes from Minimal Approximation Error to Principal Components.
5. S. Duffner and C. Garcia. Face recognition using non-linear image reconstruction. In International Conference on Advanced Video and Signal- Based Surveillance (AVSS), London, UK, September 2007.
6. Dr. H.B. Kekre, Sudeep D. Thepade, Tejas Chopra "Face and Gender Recognition Using Principal Component Analysis" (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010, 959-964.
7. Abhishek Kumar, Deepak Gupta, Karan Rawat "An Advanced Approach of Face alignment for Gender recognition Using PCA" SPRINGER LINK (AISC SERIES) SOCPROS ISSN NO-2194-5357 December 2012
9. A.A. Mohammed, R. Minhas, Q.M. Jonathan Wu, M.A. Sid-Ahmed Evaluation of face recognition technique using PCA, wavelets and NAIVE BAYES CLASSIFIER Pattern Recognition, Volume 44, Issues 10–11 Elsevier, 2011.
10. Wankou Yang Laplacian bidirectional PCA for face recognition, Neuro computing, Volume 74, Elsevier, 2010.